DOCUMENT RESUME

ED 259 021 TM 850 391

AUTHOR Varnhagen, Stanley; Calder, Peter W.
TITLE A General Testing Model for Microcomputer Assessment in Education.
PUB DATE Apr 85
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Computer Software; Educational Testing; Elementary Secondary Education; Foreign Countries; Individual Testing; *Microcomputers; Objective Tests; Teacher Role; *Test Construction; *Test Format; Testing Problems; *Testing Programs; Test Items; Test Use; Verbal Tests
IDENTIFIERS Canada; Microcomputer Diagnostic Testing Project (Canada)

ABSTRACT

The Microcomputer Diagnostic Testing Project (MDTP) was designed to broaden the use of microcomputers in student testing, for elementary and secondary school use, and for a variety of test item formats. MDTP was designed to administer tests, to allow teachers to create or revise tests, to print out test forms when desired, to score tests, to perform item analysis, and to be user-friendly. The first year of the three-year project resulted in the development of multiple choice, true-false, and A or B test items. System programs were written in BASIC for Apple microcomputers. Field testing classrooms were selected from several school districts in North-Central Alberta, Canada. Feedback was obtained and modifications were made as the project continued. There were problems associated with the way computers were distribued in schools, and the way in which various types of test formats were administered. It was found that this testing program was more readily used at the elementary level. It was also discovered that a simple way to incorporate mathematical characters was needed. Teachers' attitudes, based on preliminary surveys, were positive, but not overwhelmingly positive. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED259021

A General Testing Model for Microcomputer Assessment in Education

BEST COPY AVAILABLE

Stanley Varnhagen

Peter W. Calder
Department of Educational Psychology
University of Alberta
Edmonton, Alberta, Canada T6G 2G5

Paper presented at the annual meeting of the American Educational Research Association, April 1985, Chicago, Illinois.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Varnhagen, Stanley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ABSTRACT

The bulk of microcomputer testing literature has been concentrated in item banking and computer adaptive testing systems. Common to both formats is use of a multiple choice format and a bias towards upper-secondary and college age students. A more general model of testing is presented to help broaden the use of microcomputers for testing. Programs are being developed and examined for each of the components in a school setting. Before microcomputers can be widely accepted in schools problems with logistics, flexibility of program format, flexibility of testing format, and independence of testing components must be overcome.

A General Testing Model for Microcomputer Assessment in Education

As the number of microcomputers increases throughout education, so do the different uses to which the microcomputer is put. Among its facilities, the microcomputer offers the potential for individualized testing, automatic scoring, and randomized question ordering. However, if it is the desire of education to have microcomputers to take on this new role for the long run, then care must be taken in developing a model that is properly suited to the needs of education.

This paper presents a general model for assessment in education using microcomputers and discusses the initial microcomputer programs designed to use the model. We have developed the model into the Microcomputer Diagnostic Testing Project (MDTP). The goal of the model associated with MDTP is to provide a workable framework for the development of a series of microcomputer programs that present different types of test formats, such as multiple-choice, fill-in-the-blank, and oral presentation formats. In addition, MDTP allows the teacher to use as many or as few of the different components as desired. Finally, this project attempts to demonstrate the way in which such a system can gain widespread acceptance throughout different grade levels and subject areas.

Several interesting trends emerge from an examination of the literature on microcomputers and testing. The bulk of the literature seems to be concentrated in two specific areas. The first area is in developing item banking programs where the user has a choice of questions for each test item from a larger pool of items (cf. Hsu & Nitko, 1984; Johnson, Willis, Seely, & Moore,

BEST COPY

1981). This allows several forms of the same test to be given, minimizing the risk of cheating and allowing retesting on the same material. Item bank programs are typically used with college or upper secondary level students where the demand on the instructor is often quite high due to the large number of students that will take a test at the same time and the repetitiveness of the material being taught over time. The second area of research concentration is computer adaptive testing systems (cf. McKinley & Reckase, 1980; Vale, 1981). With this form of testing, the goal is to accurately assess the area being measured with fewer items and in less time. Typically, with computer adaptive tests, each item is chosen based on the student's performance on previous items. Most of the computer adaptive testing research is limited to college age students and in the military where aptitude tests are of high importance. Because much of the testing research has been concentrated to item banking and computer adaptive testing, most of the research has been conducted with college level and upper secondary students.

A common factor of these and other microcomputer testing programs is the multiple choice test format. A great deal of literature exists on statistical methods for testing item validity and reliability within the multiple choice testing format (cf. Thorndike, 1971); the computer is well suited to perform these operations. This analytic ability, plus the simple use of the computer's keyboard makes microcomputers well suited for this type of testing.

This dominance of the multiple choice format may not cor-

BEST COPY

respond to the frequency of the general format in education as a whole, however; especially at the elementary level or non-science subject areas where subjects such as spelling (cf. Varnhagen, 1984) are tested. While computers are well suited to the needs of item banking and adaptive testing, the needs of education itself across grade levels and subject areas should be carefully examined so microcomputers can best be used. For example, if the short answer testing format is found to be heavily used at the elementary school level, computer testing programs should be written accordingly.

The MDTP model was originally developed to demonstrate the following functions:

I) Catalog and briefly describe the tests that are included in the system.
   A) Catalog and describe other tests that are not part of the system (including written, non-computer tests).

II) Create new tests for the system or revise existing items by the teacher or other users. This feature allows the entry/revision of new tests or tests previously given in written form.

III) Administer the test on computer to students on an individual basis. Students are able to take any test previously entered on the MDTP system.

IV) Print out tests entered into the system, included tests to be printed out for written administration only. With this option system tests can printed out and given to the class in written form; for example, when facilities do not permit individualized computer testing.

V) Store the results for later retrieval either after a student has taken the test on the computer or if added separately by either the teacher or from a card reader.

VI) An option for reading mark-sense paper and then treating the data in the same manner as data from tests administered on computer.

BEST COPY

VII) Reliably score the results of tests from preset criterion and perform a limited interpretation of the data based on preset criteria previously entered by the teacher (e.g., 92% or higher = "A", 82% - 91% = "B", etc).

VIII) Print out various types of reports based on the data, including:
A) Student scores by class list.
B) Item analysis of the different questions.

IX) Maintain system flexibility so that the programs can be continually expanded to incorporate additional features or modified for specific applications.
A) For example, the ability to drive a cassette recorder would allow the system to administer prerecorded oral spelling tests or give other types of verbal tests and instruction.

X) Finally, the system attempts to be simple and sufficiently standardized to be used effectively by computer-naive teachers and students. This includes having consistent conventions throughout the different programs.

The MDTP system has recently been implemented in the schools. It is continually being evaluated by the teachers using the system. Modifications are made to the programs and to the general model based on these comments and suggestions. Our goal is to develop a system that teachers will use and find easy to incorporate into their classroom testing. The specific benefit to the teacher and student is continually being examined. In addition, care is being taken to not have the computer used in situations where existing methods are easier or more beneficial. Finally, the effect on the students of the system is being closely examined. We have not assumed that the student will necessarily perform the same on or off computers.

The first year (April 1984-August 1985) of the three year project has involved developing specific components of the programs and finalizing the model itself. Currently, forced-choice (i.e., multiple choice, true-false, and A or B) tests are the

BEST COPY

only testing format operating. Programs are written in BASIC for Apple microcomputers.

Field test classes were selected from several school districts in North-central Alberta, Canada. The classes include different grade levels and subject areas, including French immersion. Teachers agreed to use the programs in conjunction with normal testing in their classroom. The teachers were instructed to attempt to identify testing applications that are appropriate for the programs, but not to force their tests to fit the required testing format.

During this first year of the project, most of the time was spent developing the system components, obtaining feedback from teachers who use the system, and field testing the programs. Emphasis was placed on the creation of the test development programs themselves, with program additions and modifications made in response to teacher suggestions. These continual modifications were made at the temporary cost of being able to develop all the necessary programs for the model. The components that have been delayed include the test description, limited interpretation, and item analysis components of the MDTP model.

Although the model has provided an adequate general framework for the development of the system programs, several specific requirements have emerged and are important to take into account in the development of testing programs. In addition, some of the problems that have been encountered during the project may be potentially found in other microcomputer testing related situations. A fundamental difficulty concerns the way computers are distributed in many schools. Either computers are distributed

BEST COPY 8

across the entire school with, at best, two computers per classroom or the computers are concentrated in one or two locations in microcomputer labs. If computers are distributed across the school, then the testing must be done individually and across a long period of time. This often is in contrast to the way tests are typically given, in which the entire class takes a test at a specific time. On the other hand, if the computers are concentrated in a lab setting, then testing can be accomplished in one or two sittings; however, enough copies of the program and the computer-administered test must be available for each machine. There is no guarantee that the computer distribution in a particular school will match the way in which the teachers in that school generally administer their tests.

Related to this fundamental problem, it was found that a paradox exists between the <u>type</u> of tests and the <u>way</u> in which the tests are administered. Elementary classes are probably best suited to individualized presentation of tests but, on the whole, elementary school teachers use multiple choice tests the least. On the other hand, multiple choice tests increase in importance throughout secondary education as class size increases, but individualization becomes increasingly impractical due, among other factors, to the increased need to test an entire class at one time. It then becomes more efficient to use this system simply as a means to write the tests, have the tests printed out, given in the normal paper and pencil way, and then hand scored. In the future, when the card reading facility is available, the results can be read back into the computer.

BEST COPY

Regardless, so far we have found that our program is more readily used at the elementary level. This is likely due to the following reasons: First, logistically, computers and classes are often suited to individualized assessment in the elementary school setting. That is, elementary classes are often broken into various groupings that work on the same subjects at different paces, so testing may occur in small groups or on an individualized basis. Second, elementary teachers as a rule perform more informal assessments of their classes. That is, elementary teachers are generally likely to give many short quizzes in the various subject areas taught. Finally, elementary tests are often simpler and shorter than those required at the secondary level. That is, both the length of the test questions, and the number of questions on a typical test will generally be shorter at the elementary school level. It had not been our original intent to develop our system to favor any level of education, but due to these major factors and others, our system has initially been found to be best suited for testing at the elementary level.

Another problem is being able to deal with special characters (such as the division sign), illustrations, and figures. It is possible to modify a computer program to allow for special characters and figures, but the relative simplicity of the program is then sacrificed. This is the case because it is necessary for the user to learn additional procedures to choose the appropriate character sets in addition to the other standard question entry procedures. Still, because of the large demand for mathematical characters, a current programming priority is to

BEST COPY 10

allow for the addition of special characters.

The data sources used to evaluate the program and the model come from the ability to write and modify the programs according to explicit objectives and continual formative and summative evaluations performed in the classroom. In addition, teachers and students are being given surveys for their comments and opinions. One of the major means of evaluating the system concerns the teachers' interest in continued work with the programs after the initial evaluations and interest of other teachers to become involved in the project.

To this point, there has been continued teacher interest in the program. Data is being continually collected as the field tests proceed. Very preliminary teacher survey data indicates a positive but not overwhelming, positive attitude toward the MDTP programs. The teachers see the system as comparable to the procedure for creating a written test in terms of the amount of time required. They also express an interest in continued use of the programs, especially if changes they have suggested are made and if the program remains available for an entire semester. The teachers felt they were at a particular disadvantage in trying to use this system in the middle of the semester.

Evaluation data is continually being collected and modifications made accordingly. This interaction between teachers and the program developers has led to the development of important features that allow greater generality of the MDTP system. One of the original suggestions made by a teacher examining the program was to include an option to allow students to skip the

BEST COPY

question being presented. After all the questions were presented they would be re-presented for review and the student would have the option of changing his or her answer. Since this is possible with most written tests, it was decided to incorporate this option, in the MDTP system as the skip/review option.

When skip/review is used, the student is presented with the choice of skipping the question in addition to the different answer options. After the initial presentation of the questions, each question is shown a second time. The student has the choice of doing nothing, accepting his or her answer, or changing the answer. At the end of this and subsequent presentations, the student is asked if he or she wants to review the questions and answers yet another time.

In order to examine whether and how the feature is used by the student, MDTP records on disk the student's answers from both before and after the the review procedure, keeping track of items the student skipped and changes the student made to his or her answers. Through comparing the files, the use of the skip option and the ability to change answers can be easily noted.

Preliminary data has shown that about half of the students have not taken advantage of this feature; in other words, they have not used the skip option or changed their answer during review. For those students who do use the feature, their overall test score tends to either remain the same or improve through using the skip option. When a student skips a problem, he or she tends to subsequently get the problem correct about 50 percent of the time, which is above a chance level.

If this finding is upheld with subsequent data, then the

importance of allowing skipping and reviewing will be supported. In addition, these results would imply that non-inclusion of this feature in a computerized multiple choice program may hurt the scores of some students. Given this knowledge, it may be possible to challenge the comparability of other computerized testing programs and a normal paper and pencil test where skipping and review can be a matter of course.

The awareness of the potential aid that microcomputers can offer to student assessment is becoming increasingly better understood. Programs now exist that allow for item banking, adaptive testing, creation of simple multiple choice tests, and aid in analysis and recording of results. However, currently most of the systems are separate and disjoint. Those systems that are more general often have several restrictions in format and cannot incorporate deviations. The MDTP system, on the other hand, is flexible to the needs of teacher, can be used in a component manner or as an entire system, and is employed in several grades and subject matters. It should be relatively simple to modify the existing programs in response to feedback we have received from our field testing experience and surveys. For example, we hope to learn which features were most important and useful to the teachers so we know where to concentrate our own effort. This information in turn is helping us better understand the testing process itself. We hope to continue to develop MDTP according to our model, and, if it gains widespread acceptability, then we will have demonstrated an effective use of microcomputers in education.

References

Hsu, T. & Nitko, A. J. (1984) A microcomputer item bank for use in classroom testing. Paper presented at annual meeting of the American Educational Research Association, New Orleans.

Johnson, K. J., Willis, W. V., Seely, D., & Moore, J. W. (1981) Computer-assisted test construction in chemistry. Journal of Chemical Education, 58, 177-181.

McKinley, R. L. & Reckase, M. D. (1980) Computer applications to ability testing, AEDS Journal, 13, 193-203.

Thorndike, R. L. (Ed.)(1971) Educational measurement, 2nd Ed. Washington D.C. American Council of Education.

Vale, C. D. (1981) Design and implementation of a microcomputer based adaptive testing system. Behavior Research Methods and Instrumentation, 13, 399-406.

Varnhagen, S. (1984) Use of feedback on microcomputers to teach spelling in elementary school. Unpublished doctoral dissertation, University of California, Santa Barbara.

BEST COPY